Countering Disinformation, Hate Speech and Incitement to Violence on Social Media During Elections

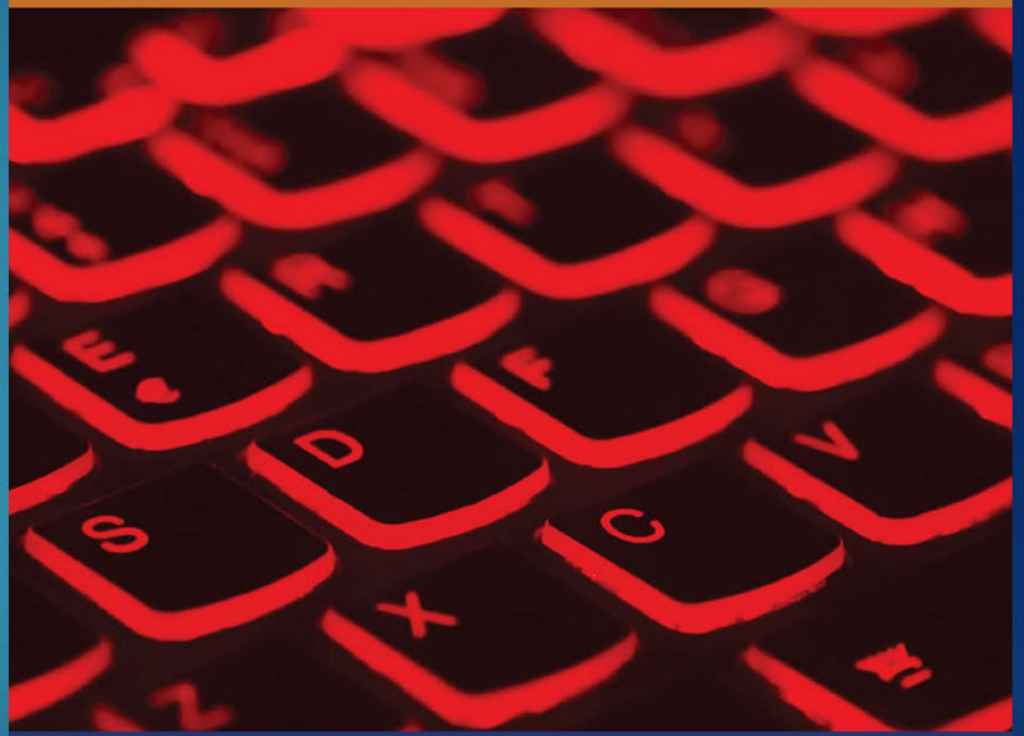
Countering Hate Speech in Elections: Strategies for Electoral Management Bodies

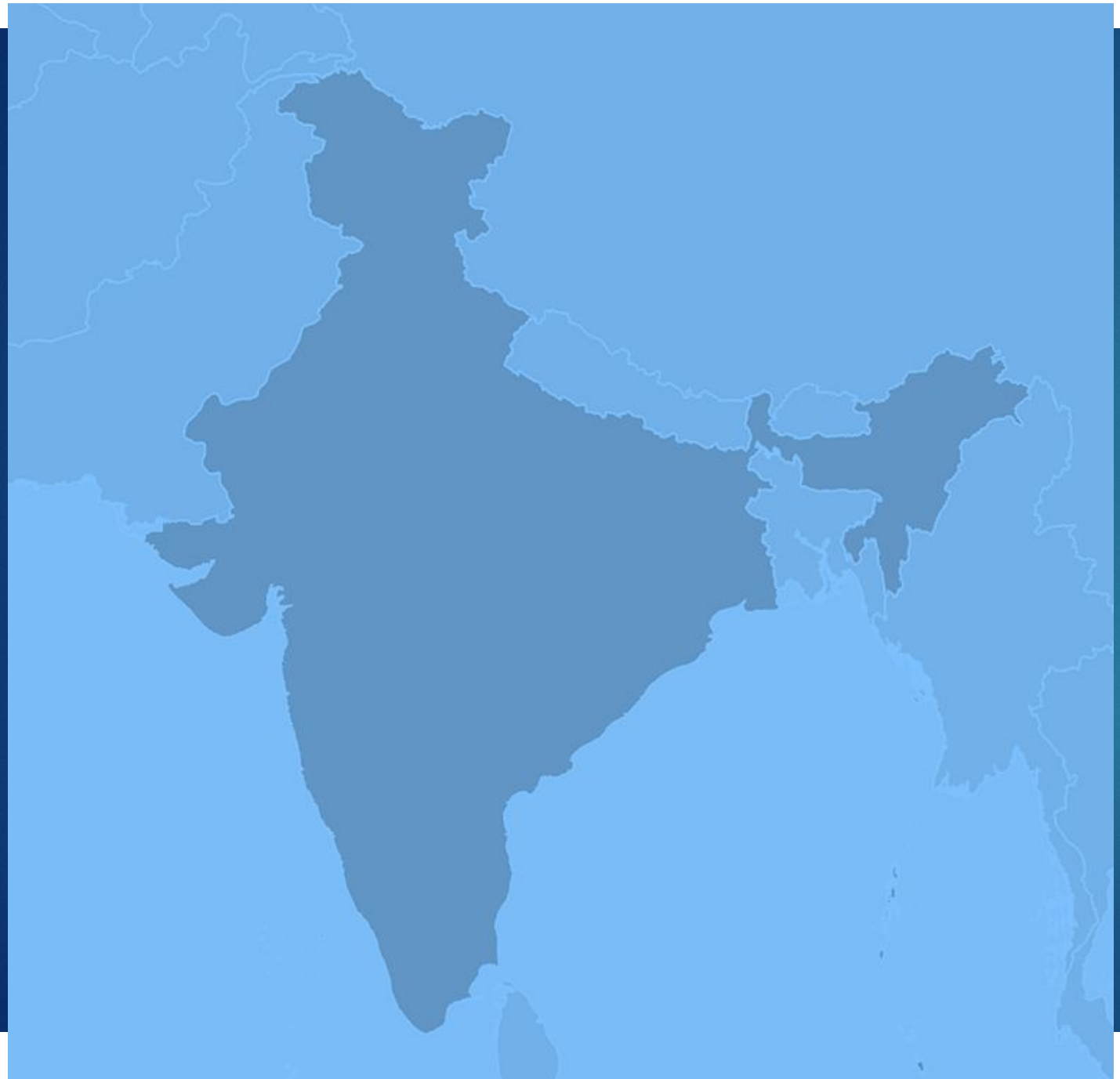
IFES White Paper



Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions

April 2019





India

Promoting Social Cohesion and Inclusion in Social Media



INTERNATIONAL FOUNDATION FOR ELECTORAL SYSTEMS · IFES · 30 Years ·

IFES FACEBOOK PARTNERSHIP

- 6 Months
- 8 Workshops
- 150 NGOs in 12 states and a dozen national networks
- Cross section of targets – women, LGBTIQ, minorities, indigenous (tribal), marginalized castes, people with disabilities, youth, artists, media
- Facebook Trusted Partner Program
- Counterspeech Practice & Effectiveness – Collaboration with GW
- Functioning effectively in closing spaces

Countering Hate Speech in Elections: Strategies for Electoral Management Bodies

IFES White Paper



Responding to Hate Speech & Disinformation

- No Response – Freedom of Speech
- Statutory Responses
- Non-Statutory Responses

What can EMBs & Electoral Stakeholders do to Counter Hate Speech, Hate Spin & Disinformation?

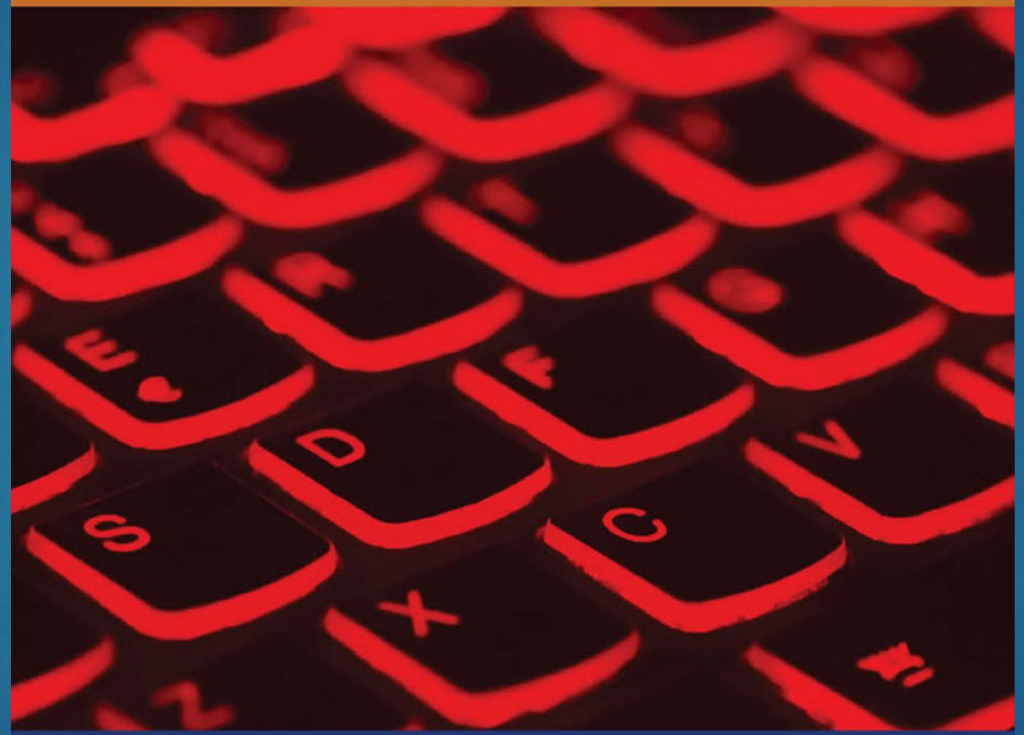
- Diverse stakeholder engagement
- Speak out early and often against discrimination & Hate Speech
- Create open spaces for pluralistic public dialogue
- Develop media literacy and critical thinking skills

What can EMBs & Electoral Stakeholders do to Counter Hate Speech?

- Adjudicate effectively and responsibly
- Train electoral stakeholders
- Raise awareness and educate voters
- Advocate for a better legal framework to combat hate speech
- Develop partnerships with social media platforms

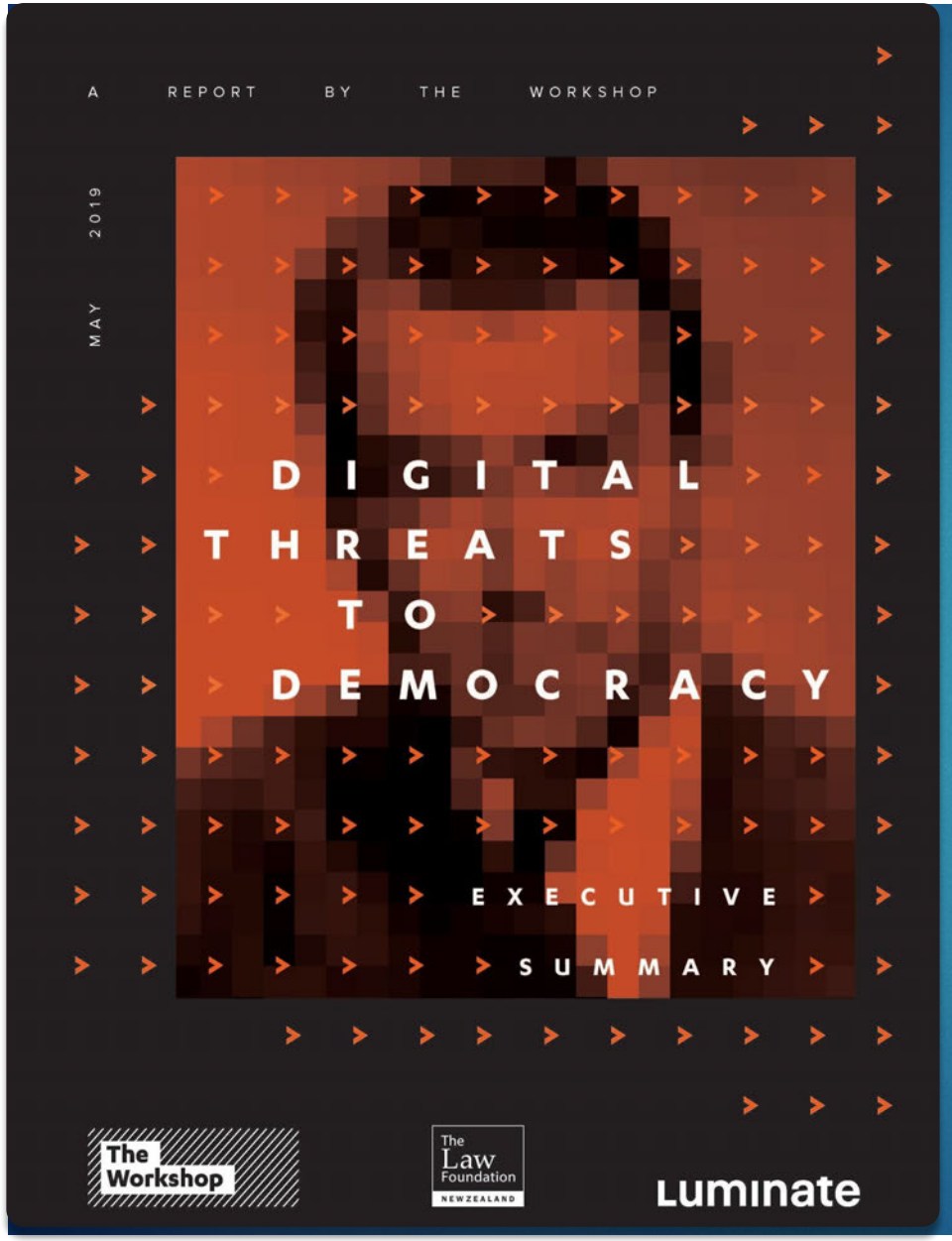
Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions

April 2019



	TARGETING THE ACTOR	TARGETING THE MESSAGE	TARGETING THE MODE OF DISINFORMATION	TARGETING THE INTERPRETER	MITIGATING RISK
Hate Speech	Review of the legal framework for identifying and sanctioning producers of hate speech. Development of more effective investigation and enforcement tactics to prosecute offenders within the bounds of the law.	Empower public actors, such as EMBs, to model inclusive practices and speak out against hate speech. Provide alternative narratives by elevating the voices of marginalized groups.	Codes of conduct for media actors to prevent the spread of hate speech through traditional media sources. Pressure on social media companies to slow the spread of viral hate speech, or enforce their hate speech community standards and remove hateful content.	Civic education that promotes tolerance. Awareness campaigns that educate the public about hate speech.	Coordinated security planning by multiple agencies to mitigate the risk of electoral violence.

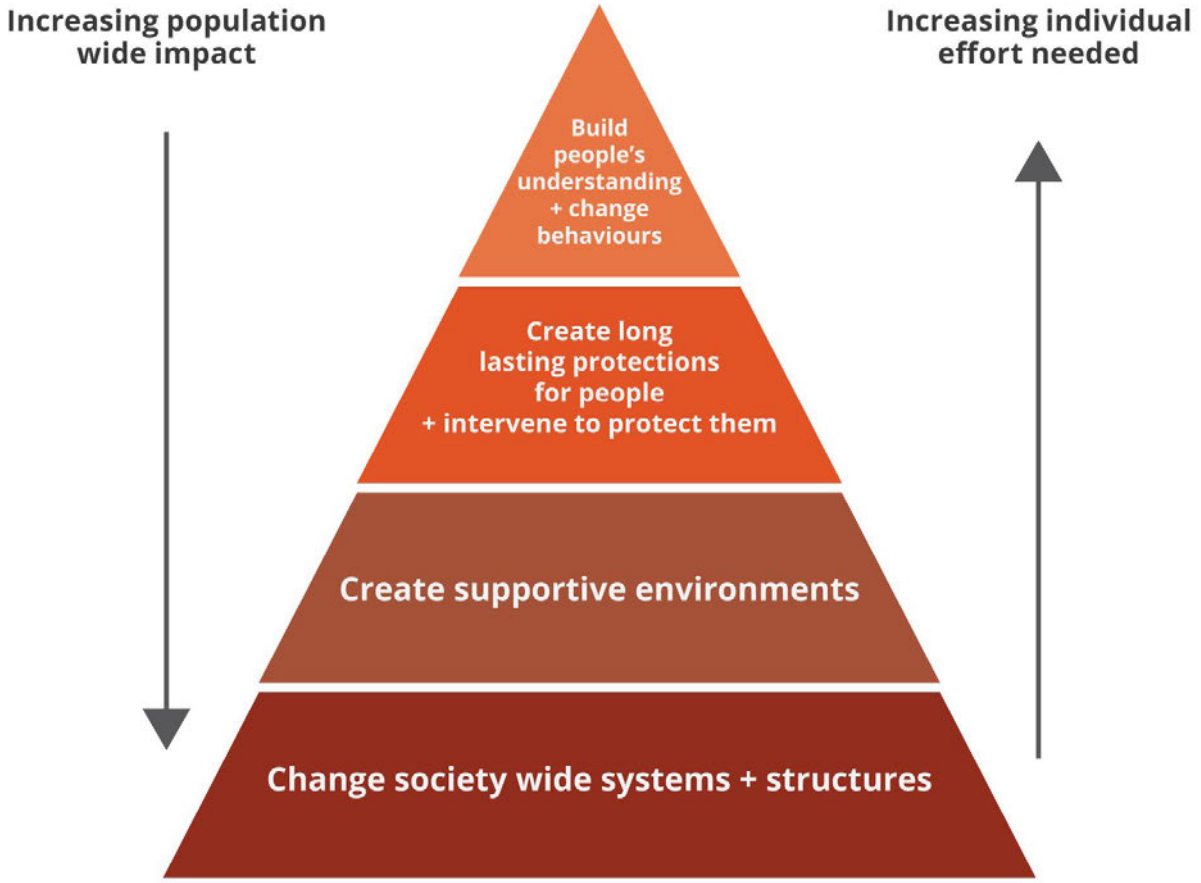
	TARGETING THE ACTOR	TARGETING THE MESSAGE	TARGETING THE MODE OF DISINFORMATION	TARGETING THE INTERPRETER	MITIGATING RISK
Disinformation	Political finance regulation to prohibit domestic political actors from engaging in the production and dissemination of inauthentic content.	Advocacy to demand greater transparency from social media companies on inauthentic content circulating on their platforms. Research and monitoring of evolving trends in inauthentic content production.	Identification and removal of inauthetic coordinated behavior on social media platforms.	Civic education that promotes media literacy. Public service announcements that educate the public about disinformation campaigns.	Crisis planning for EMBs on how to address and mitigate the impacts of disinformation during critical parts of the electoral cycle.



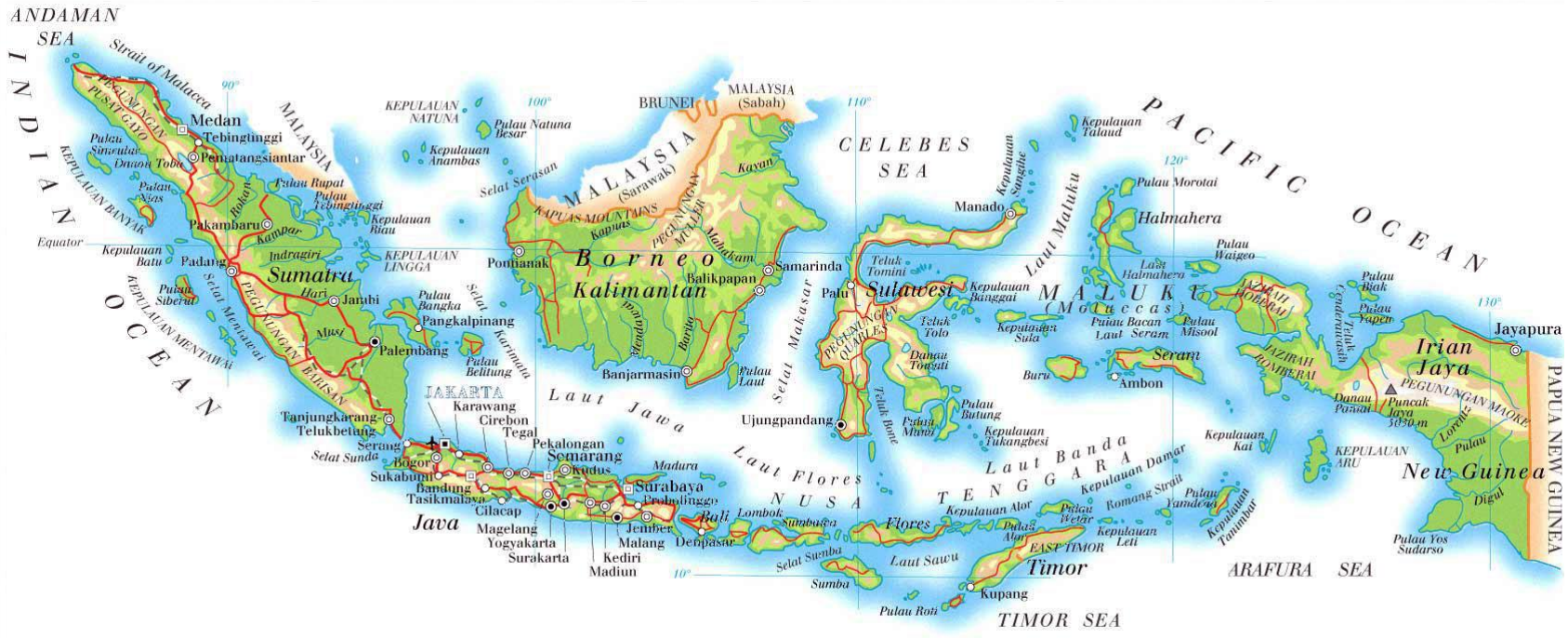
Three Core Problems:

1. Platform Monopoly
2. Algorithmic Opacity
3. Attention Economy

Wellbeing Impact Pyramid



Adapted by The Workshop from Frieden (2010)


ANDAMAN SEA
INDIAN OCEAN
Strait of Malacca
MALAYSIA
BRUNEI
MALAYSIA (Sabah)
MALAYSIA (Sarawak)
CELEBES SEA
PACIFIC OCEAN
Laut Maluku
Halmahera
Manado
Medan
Tebingtinggi
Pematangsiantar
Pakambaru
Padang
Sumatra
Jambi
Palembang
Pangkalpinang
Pontianak
Borneo
Kalimantan
Banjarmasin
Balikpapan
Samarinda
Palu
Sulawesi
Ujungpandang
MALUKU (Moluccas)
Ambon
Seram
Buru
Irian Jaya
New Guinea
Jayapura
PAPUA NEW GUINEA
JAKARTA
Karawang
Cirebon
Tegal
Pekalongan
Semarang
Kudus
Surabaya
Tanjungkarang
Telukbetung
Serang
Bogor
Sukabumi
Bandung
Tasikmalaya
Cilacap
Magelang
Yogyakarta
Surakarta
Jember
Malang
Kediri
Madiun
Java
Bali
Denpasar
Lombok
Sumbawa
Flores
Timor
EAST TIMOR
Kupang
Laut Sawu
Laut Jawa
Laut Flores
Laut Banda
NUSA TENGGARA
ARAFURA SEA
TIMOR SEA
Selat Makasar
Selat Sunda
Selat Malaka
Selat Sumba
Equator
Pulau Natuna Besar
Kepulauan Anambas
Kepulauan Riau
Pulau Rupat
Pulau Bangka
Pulau Belitung
Kepulauan Lingga
Kepulauan Batu
Pulau Nias
Pulau Simeulue
Kepulauan Banyak
Pulau Siberut
Kepulauan Mentawai
Kepulauan Talaud
Kepulauan Sangihe
Pulau Morotai
Pulau Waigeo
Kepulauan Sula
Kepulauan Banggai
Pulau Taliabu
Pulau Obi
Pulau Bacan
Pulau Misool
Kepulauan Tukangbesi
Pulau Butung
Kepulauan Kai
Kepulauan Aru
Kepulauan Damar
Kepulauan Alor
Kepulauan Leti
Kepulauan Tanimbar
Pulau Yos Sudarso
Pulau Roti
Pulau Sumba
Pulau Laut
Pulau Bawean
Pulau Madura
Kepulauan Karimata
Pegunungan Jayawijaya
Puncak Jaya 5030 m
Danau Paniai
Danau Tondano
Danau Matano
Teluk Tomini
Teluk Tolo
Teluk Bone
Teluk Cenderawasih
Kapuas Mountains
Pegunungan Meratus
Pegunungan Quarles
Pegunungan Maoke
Pegunungan Barisan
Pegunungan Muller
Pegunungan Kapuas
Kayan
Mahakam
Barito
Musi
Hari
Indragiri
Kampar
Rokan
Kepulauan Natuna
Kepulauan Lingga
Selat Serasan
Selat Karimata
Selat Berhala
Selat Malaka
Selat Sunda
Selat Sumba
Selat Makasar
Laut Halmahera
Romang Strait
Pulau Wetar
Pulau Alor
Pulau Yamdena
Digul
Lorentz
Pulau Yapen
Pulau Biak
Kepulauan Raja Ampat
Kepulauan Widi
Kepulauan Banda
Kepulauan Sula
Kepulauan Obi
Kepulauan Bacan
Kepulauan Seram
Kepulauan Misool
Kepulauan Aru
Kepulauan Tanimbar
Kepulauan Leti
Kepulauan Alor
Kepulauan Damar
Kepulauan Tukangbesi
Kepulauan Banggai
Kepulauan Sula
Kepulauan Talaud
Kepulauan Sangihe
Kepulauan Natuna
Kepulauan Anambas
Kepulauan Riau
Kepulauan Batu
Kepulauan Nias
Kepulauan Simeulue
Kepulauan Banyak
Kepulauan Mentawai
Pulau Siberut
Pulau Nias
Pulau Simeulue
Pulau Rupat
Pulau Tebingtinggi
Pulau Bangka
Pulau Belitung
Pulau Laut
Pulau Bawean
Pulau Madura
Pulau Natuna Besar
Pulau Morotai
Pulau Waigeo
Pulau Biak
Pulau Yapen
Pulau Taliabu
Pulau Obi
Pulau Bacan
Pulau Misool
Pulau Butung
Pulau Wetar
Pulau Alor
Pulau Yamdena
Pulau Yos Sudarso
Pulau Roti
Pulau Sumba
Pulau Digul
Pulau Lorentz

The Race Against SARA and Hoaxes in Indonesian Elections

In Bahasa Indonesia "suku, agama, ras, dan antargolongan," or ethnicity, religion, race and inter-group relations. "Berita hoax" refers to disinformation.

STRATEGY OF ILLIBERAL POLITICAL LEADERS

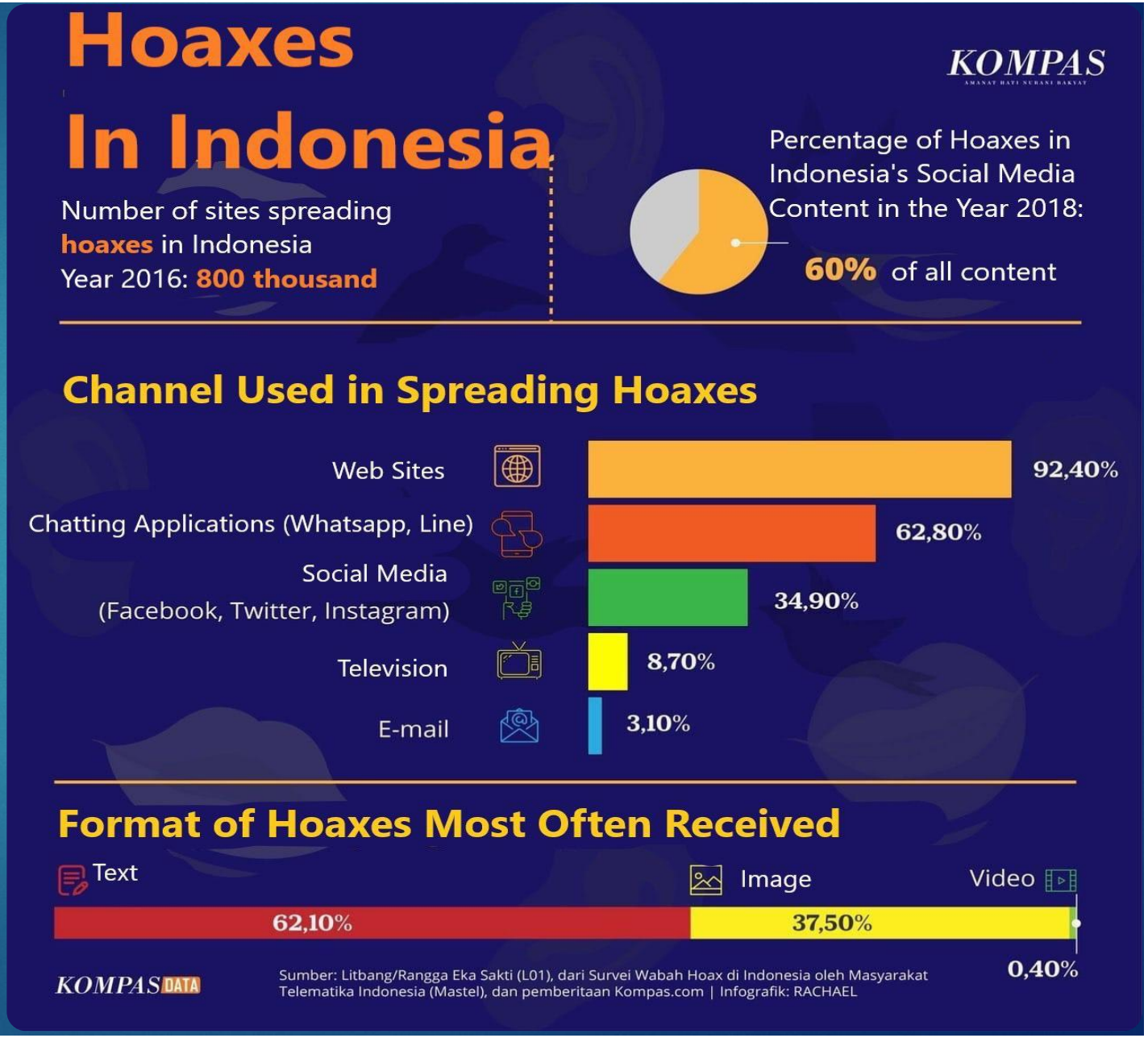
- Rather than shutting down dissenting voices harness the democratizing power of social media —jamming signals, sowing confusion.
- No longer need to silence the dissident shouting in the streets; use a “megaphone” to drown dissent out.
- “Censorship through noise”

“The 2020 Disinformation War.” MCKAY COPPINS staff writer at The Atlantic and author of The Wilderness, a book about the battle over the future of the Republican Party.

Indonesia – A Successful Example

Experts worried about domestic [US] disinformation are looking to other countries for lessons. The most successful recent example may be **Indonesia**, which cracked down on the problem after a wave of viral lies and conspiracy theories pushed by hardline Islamists led to the defeat of a popular Christian Chinese candidate for governor in 2016. To prevent a similar disruption in last year's presidential election, a coalition of journalists from more than two dozen top Indonesian news outlets worked together to identify and debunk hoaxes before they gained traction online.

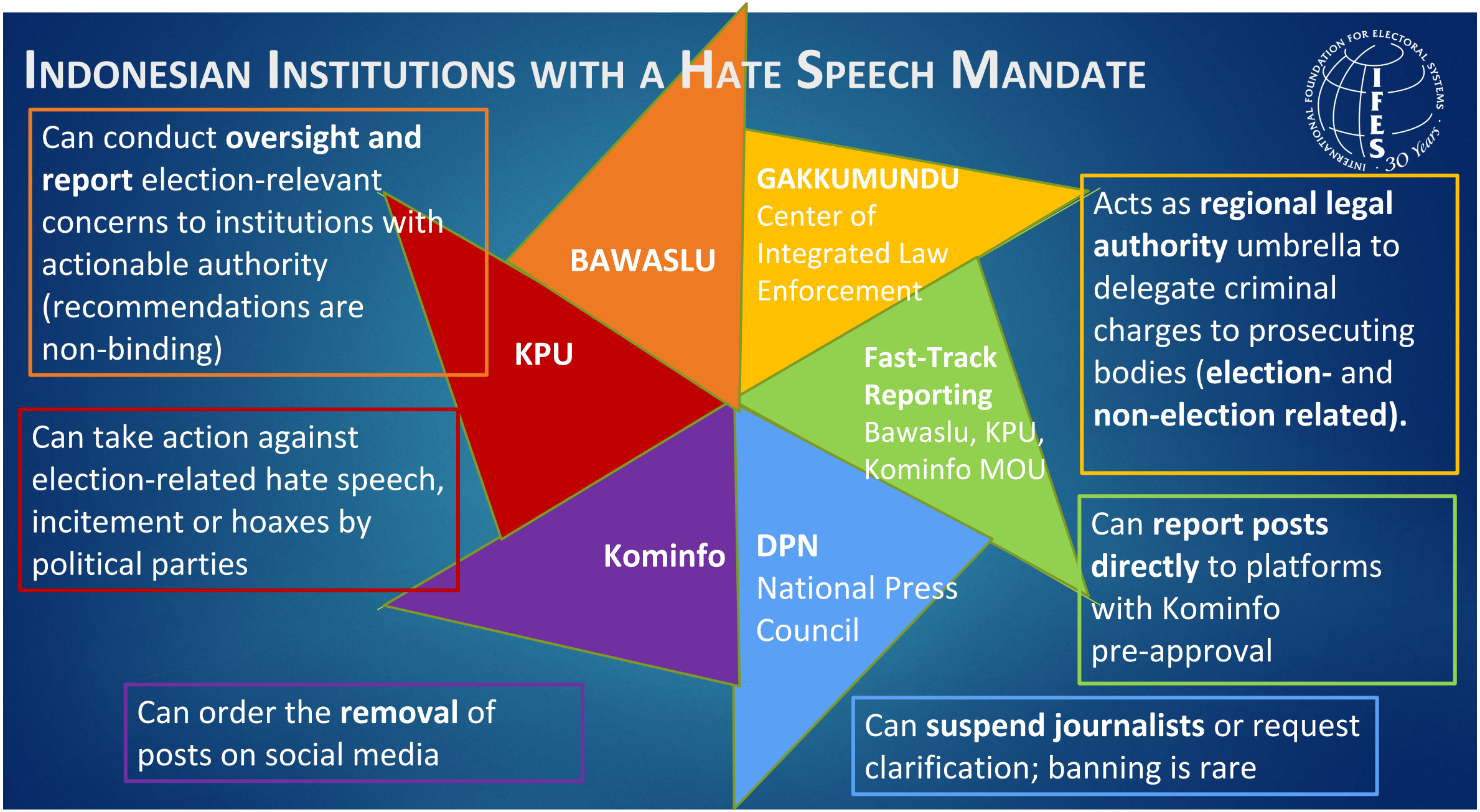
INDONESIA


Hoaxes
In Indonesia
KOMPAS
Number of sites spreading hoaxes in Indonesia
Year 2016: 800 thousand
Percentage of Hoaxes in Indonesia's Social Media Content in the Year 2018:
60% of all content
Channel Used in Spreading Hoaxes
Web Sites
92,40%
Chatting Applications (Whatsapp, Line)
62,80%
Social Media (Facebook, Twitter, Instagram)
34,90%
Television
8,70%
E-mail
3,10%
Format of Hoaxes Most Often Received
Text
Image
Video
62,10%
37,50%
0,40%
KOMPAS DATA
Sumber: Litbang/Rangga Eka Sakti (L01), dari Survei Wabah Hoax di Indonesia oleh Masyarakat Telematika Indonesia (Mastel), dan pemberitaan Kompas.com | Infografik: RACHAEL

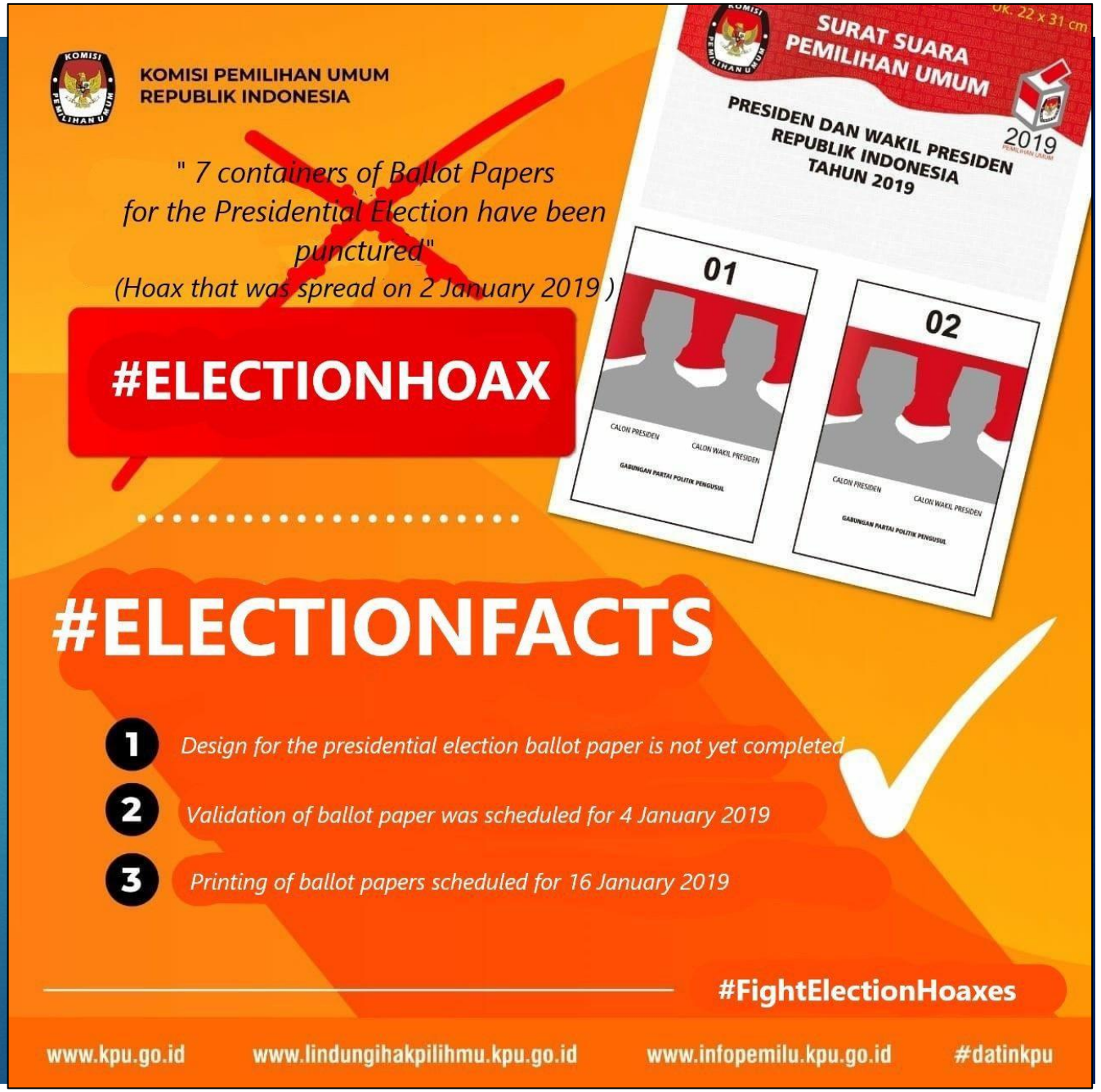
Key actors

- **Komisi Pemilihan Umum (KPU)**
- **Bawaslu**
- **Civil society organizations**, like the Indonesian Anti-Hoax Community (MAFINDO), Observation Organizations, Perludem, Religious Groups
- **Journalist/media** outlets
- **Government ministries**
- **Social Media Platforms**

Indonesian Institutions with a Hate Speech Mandate



✔ KPU LAUNCHES MYTH BUSTER CAMPAIGN TO COUNTER DISINFORMATION



Fact-Checking and Hoax Buster Tools



MAFINDO launches Hoax Crisis Center in Pontianak, West Kalimantan



Public Awareness Campaigns to Counter Hate Speech & Misinformation



MAFINDO hosted an information booth & distributed brochures and other materials. Campaign covered the perils of hoaxes & shared strategies to stop their spread Encouraged public to fact-check before sharing information

TurnBackHoax platform and Hoax Buster Tools application.

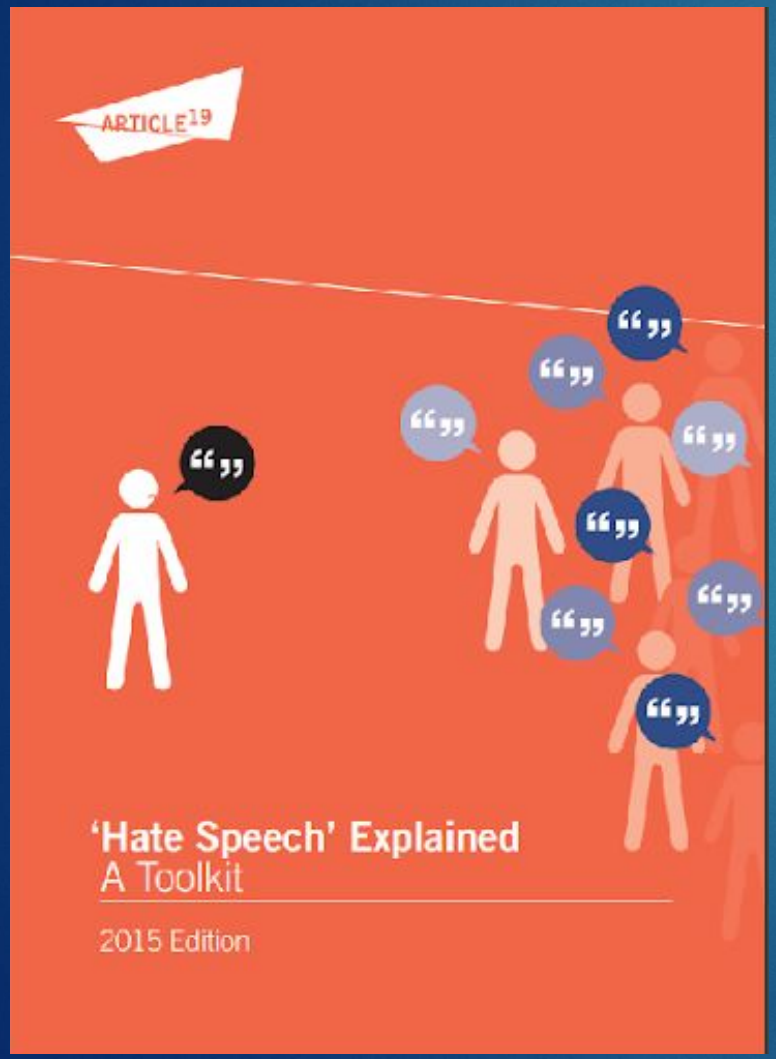
Multi-Stakeholder Declaration against Hoaxes

- Bawaslu-led declaration: ***Reject and Counter Vote Buying, Insults, Incitements, and Divisive Conflict*** in the 2018 Regional Election and 2019 General Election
- **102 signatories**, including CSOs, religious organizations, youth groups, and new media platforms such as Facebook, Twitter and Google.

Declaration includes commitments to:

- ✔ Reject intimidation and disinformation;
- ✔ Promote peaceful journalism; and
- ✔ Reject communal incitement

CSO-Led Training of Trainers on Countering Election-Related Incitement





EMBs & CSOs CREATE PSAs



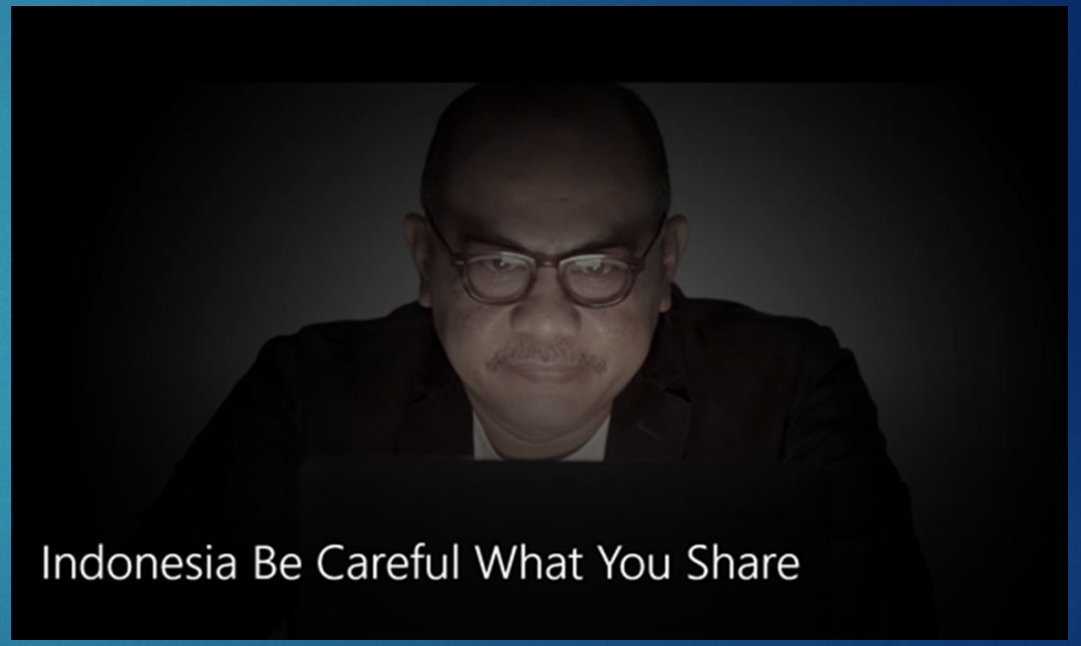
PSAs highlight ***positive stories and messages to deliver unique calls to action to counter incitement of violence, hate speech and hoaxes, and promote media literacy and inclusive participation***.

IFES hosts co-creation workshops for PSA development

Countering Hate Speech and Disinformation: PSAs



MAFINDO PSA: *Verify*



Bawaslu PSA: *Campaign Against Incitement*

Social Media Platforms

- Facebook/WhatsApp, Twitter, Google worked in partnership with trusted CSO partners to remove harmful content and promote PSAs and Countering Disinformation
- Platform-level responses – WhatsApp reduced number of forwards (5) and number of members (256) per group

IMPACT 2018

Institute for Policy Analysis of Conflict (IPAC) in its 2018 “Update on Local Election Result in West Kalimantan and Papua” report.

> “If the predicted violence did not materialize in West Kalimantan, one reason may have been the work of a Hoax Crisis Center led by NGOs and journalists committed to investigating, clarifying and where necessary rebutting provocative messages circulating online or on mobile phone texts. Tensions were also defused by the loser’s gracious acceptance of the results.”

Key Elements of the Intervention

- Collaboration between election authorities, government entities, and civil society;
- Collaboration between civil society with different but related areas of focus and expertise;
- Involvement of religious and community leaders;
- Public service announcements and targeted local voter education efforts including digital literacy;
- Fact-checking and countering misinformation and disinformation.

8 Strategies for Counter-speech Based on Research Findings

Research Findings on Counterspeech

Who is producing hate speech? Who are the different audiences? What are the ways in which they are impacted both by hate speech and counterspeech?

1. You are more likely to impact the views of those who are vulnerable to joining hate groups than those who are already members with more deeply entrenched views. So speak to the broader audience viewing the discussion.
2. Confronting hateful speakers usually does not change their behaviour, but it can draw other counterspeakers and change the overall discussion norm.
3. Photos and videos are shared more and reach more people beyond own network so should be used as a higher proportion of total output. Images are often more persuasive than text alone e.g. memes, graphics, photos, animated gifs and videos.

Research Findings on Counterspeech

4. For comments on hate speech posts, engage in more 'constructive' counter-speech and specific policy issues is more effective than disparaging comments (e.g. name calling).
5. Tone of engagement important. Casual and sentimental tone or offers of assistance or personal stories are more likely to get engagement with extremists that antagonistic messages and those that called attention to negative consequences of hateful speech.
6. People follow online norms or what they see when they enter an online space. This can be called contagion or herding theory. If there is civility, participants are more civil. If there is incivility, then participants follow. Counterspeakers, therefore, may want to enter online spaces in groups to have more impact.

Research Findings

7. Source credibility: The person delivering the counterspeech must be credible to the audience.). Recruit people who are influential to those whose views they are trying to change or reference them (e.g. quote from prominent figures who are credible to those they want to influence). But its not always a prominent person. Authentic peers.
8. More Effective: Warnings about consequences, labeling the act clearly, empathy and affiliation and humor.

Less Effective: Hostile and aggressive tone/insults, fact checking, harassment and silencing.

Some additional Concepts & Insights

Beware of Confirmation bias & Motivated reasoning

Providing alternate explanations but not invalidating grievance

Don’t repeat a false claim while countering it

Combine empathy with humanization

Peace & Conflict Resolution Insights

- Peace is not the absence of tension but the presence of justice.
- Peacemaking is not passive. It is the act of interrupting injustice without mirroring it.
- Hate cannot drive out hate. Only love can. A thought of hatred must be destroyed by a more powerful thought of Love.
- Society is like the human body – multiple cells, infinitely diverse but they cooperate and not compete. Unity in diversity is the operative principle.

IFES Global Tools (partial list)

- Global Research & White Papers
- Crisis Simulations e.g. Hate speech and elections piloted with EMB, Media and Police in Sri Lanka, Fiji, Seoul/AWEB Washington, DC
- Multi-stakeholder consultative working groups
- Research on disinformation – tools employed and strategies used by EMBs to counter disinfo
- Digital Literacy Curricula